8213 Computer Ne rks and ISDN Systems
20(1990)December .1/5, Amsterdam, NL
H04L 12/26M

Traffic Measurement on Variable Bit Rate (VBR) Sources with Application to Charging Principles

p. 435 - 445

Biørn VEIRØ
Copenhagen Telephone Company, Noerregade 21, DK-1199 Copenhagen, Denmark

Abstract. This paper proposes a method making it possible to measure traffic from variable bit rate (VBR) sources as they occur in the emerging Asynchronous Transfer Mode (ATM) networks. The method is a kind of mixture of a sampling process known from classical communication engineering and of the Karlsson scanning method, used in the telephone networks in the Scandinavian countries and in Germany. The link to charging comes when we choose a charging principle where (at least part of) the charge is proportional to the traffic generated to the network.

Keywords. Measurement, variable bit rate, ATM, charging.

1. Introduction

Traffic measurement serves several purposes in a conventional telephone network: One is to make a general surveillance of the traffic either in order to obtain statistics for the forecasting needed for the future extensions of the network or in order to check the promised grade of service. Another one merely serves the purpose of providing the user with a charge, which (s)he considers fair. (A linear function of the traffic generated to the network would normally be considered to have this property.)

In the Scandinavian countries (and in Germany) the so-called Karlsson scanning principle, which was first described by Karlsson [1] in 1937, is used. For a telephone user the principle is the following: When a call is set up it will be connected to a scanning process which increments a counter with regular intervals, whose length only depends on the distance of the call. The scanning process used to depend on some hardware in the subscriber office and had nothing to do with the establishment of the individual call set-up processes (except that at call set-up it was ensured, that each call would be assigned to a scanning process). The principle is still the same though the mechanism that ensures it today is different. Using this method, a call whose duration is shorter than the scanning interval, will either be "lucky" in which case it experiences no scans at all and will be free of charge, or it will be unlucky in which case it will receive a scan and the call will be charged for one whole scanning interval. The number of scans is used as a measure for the duration of the call, and obviously the scanning method introduces an uncertainty in this measurement.

Biørn Veirø received his M.Sc. degree in electronics engineering from the Technical University of Denmark in 1973. After almost 10 years in KAMPSAX International, where he participated in the development of software for the European space laboratory known as SPACELAB (including simulators, interpreters and mathematical applications) he joined the R&D department in KTAS (Copenhagen Telephone Co.). He is head of operations research, whose main responsibility is teletraffic engineering. He has been active in the European research projects COST 214 and its successor COST 224 (Methods for the Performance Evaluation and Design of Synchronous and Asynchronous Multi Service Networks) as well as in RACE, where he has mainly done economic modelling. In 1989 he became the head of the Danish delegation for the CCITT SGII (network operation).

North-Holland
Computer Networks and ISDN Systems 20 (1990) 435–445

8213 Computer Networks and ISDN Systems
20(1990)December, No.1/5, Amsterdam, NL

In Section 2 we shall make some general remarks on charging principles, in Sections 3 and 4 we shall recapture some of the main features of the Karlsson scanning method and of sampling theory respectively.

2. General Remarks on Charging Principles

In the "clean" version of the Karlsson scanning principle only the counts obtained during the calls are collected, but some network operators have chosen a version of the method, where the counter is incremented as soon as the call is established in order to account for the load on the network for the establishment of the call. Other principles could be taken into account as well, but this paper will not enter the discussion of whether one or the other method is more fair to the user. Such a discussion can be found in [2] for instance. This paper will focus only on the part of the tariff, that has to do with the traffic generated to the network.

One note should be made, however: The charging principle used in the USA for instance, where the subscriber gets a record of all the calls made in the period covered by the telephone bill, is impossible when the Karlsson method is used, because no record is kept of the fate of each individual call. The strength of the method is, that it is cheap to implement, but the major problem is that it relies on the fact that there is a relationship of confidence between the network operator and the customer. If a customer has any complaints it is more or less impossible both for the customer and for the administration to prove anything, and therefore any law-suits will be very difficult to conduct.

3. Properties of the Karlsson Scanning Method

The following description of the Karlsson scanning method follows [3]. Using the scanning method we want to measure a continuous time interval starting at $t = T_1$ and ending at $t = T_2$. The number of scans received in the duration $T = T_2 - T_1$ is a discrete random variable Z, and we define $\{Z = k\}$ as the occurrence of k scans within (T_1, T_2).

We now want to confine ourselves to constant scanning intervals, because this is the way it is actually implemented in practice. The fact that the method works with random scanning intervals as well at the cost of some accuracy, is shown in detail in [3]. If we let the scanning interval be h and the integral part of T_1/h be $|T_1/h|$ we obtain

$$\begin{aligned} \{Z=0\} &= \{T_2 \leqslant |T_1/h| + h\}, \\ \{Z=k\} &= \{h|T_1/h| + kh < T_2 \leqslant h|T_1/h| + (k+1)h\} \end{aligned} \tag{3.1}$$

when $k = 1, 2, \ldots$.

We assume that the time interval $T_2 - T_1$ has the distribution function $F(t)$ and that the time until the first scan after T_1 has the distribution density function $v_h(t)$. We also assume that T_1 is independent of the scanning times, which means that we can assume $T_1 = 0$ without any restrictions. Then we get the following probabilities:

$$\begin{aligned} p\{Z=0\} &= \int_0^h v_h(t) F(t)\, \mathrm{d}t, \\ p\{Z=k\} &= \int_0^h v_h(t)\left[F(t+kh) - F(t+(k-1)h)\right] \mathrm{d}t. \end{aligned} \tag{3.2}$$

These are the general formulae applicable to any arrival process distribution and any holding time distribution. If we now restrict ourselves to Poisson arrivals, the time elapsed from arrival of the call until the next scan will be uniformly distributed in $[0, h]$:

$$v_h(t) = 1/h, \quad 0 < t \leqslant h \tag{3.3}$$

8213 Computer N[illegible]orks and ISDN Systems
20(1990)Decembe[illegible]p.1/5, Amsterdam, NL

and we get the probabilities for general holding time distribution:

$$
\begin{aligned}
p\{Z=0\} &= p_0 = \frac{1}{h}\int_0^h F(t)\,\mathrm{d}t, \\
p\{Z=k\} &= p_k = \frac{1}{h}\int_0^h [F(t+kh) - F(t+(k-1)h)]\,\mathrm{d}t.
\end{aligned}
\tag{3.4}
$$

What we have done above corresponds to adding a $[0, h]$ uniformly distributed time to the holding time described by $F(t)$ and then truncating this sum to an integral time.

By integration by parts the expectation of (3.4) becomes

$$E\{Z\} = \sum_{k=0}^{\infty} kp_k = \frac{1}{h}\int_0^{\infty} x\,\mathrm{d}F(x). \tag{3.5}$$

So we have the important, well-known result that the original continuous distribution and the transformed discrete distribution have the same mean value, which proves that the scanning principle in a sense is fair to the user.

Let the time interval be of length $T_2 = x + kh$, $k = 0, 1, 2, \ldots, 0 < x \leqslant h$. Then the measuring error becomes negative $(= -x)$ with probability $1 - x/h$ and positive $(= h - x)$ with probability x/h, and for the mean we have

$$\mu = \int_0^h (-x)\left(1 - \frac{x}{h}\right)\left(\sum_{k=0}^{\infty} \mathrm{d}F(x+kh) + \int_0^h (h-x)\frac{x}{h}\sum_{k=0}^{\infty} \mathrm{d}F(x+kh) = 0\right.$$

This is another property, that makes the Karlsson scanning fair and for the method proposed for ATM networks we wish to retain both the zero mean error property and (3.5).

4. Review of Sampling Theory

In normal textbooks in communications engineering (e.g. [7]) the proof of the sampling theorem depends on a formulation using the Dirac delta function (δ) because it has the nice property that the convolution of a function $f(t)$ with δ at time t_0 gives the value $f(t_0)$. If it was not for some mathematical problems with the interpretation of the physical implementation of sampling systems this would be very good because the normal idea with sampling is that we want to recover the sampled signal at the receiver. This is not the idea in this measurement application. First of all we only want to have a very rough picture of what the traffic was as long as the mean error in the measurement is kept at zero and the holding time distribution has the same mean value as the scanning process as was described for Karlsson scanning in Section 3.

The sampling error that we shall make with the method described below will be even bigger than in conventional sampling systems because in normal systems we try at least to make an approximation to the δ function, which is as good as possible, but in ATM systems we need to take an average value over a time period to be specified below. If we took a true sample, we would either get the maximum bandwidth available for the access line or we would get zero. In principle there is nothing wrong with this, but we face the practical problem with the samples that we have to be able to distinguish empty time slots from information carrying ones and for that we need at least to have time to analyse the header, which is not compatible with the sampling concept.

Therefore it would be nice if a proof of the sampling theorem that did not depend on the δ function could be produced and actually such a proof has already been made in [5]. This has been translated into English and is provided as Appendix A as the author does not know of any sources in English providing such a proof.

One more thing should be mentioned: we are not really interested in sampling in the normal sense of communication theory because ordinary sampling is done in order to regain the original signal (as good as possible with the given sampling frequency), but we are only interested in having some general statistical properties retained in the measurement case and only an overall picture of the signal in the charging case.

8213 Comput[...]etworks and ISDN Systems
20(1990)December, No.1/5, Amsterdam, NL

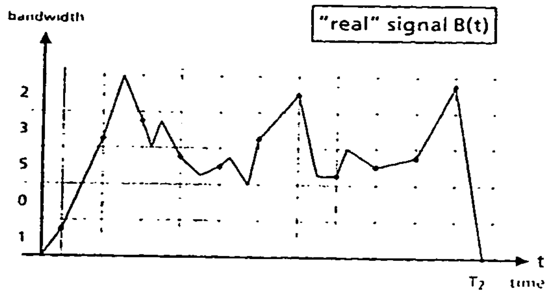

Fig. 1.

5. The ATM-measurement Method Proposed

5.1. Description of Method

The principle in the measurement method is illustrated in Fig. 1. The source starts emitting cells at time $T_1 = 0$. With regular time intervals the bandwidth function is "sampled" and the sampled value is digitised into L levels where $L \geqslant 2$. (For $L = 1$ the method degenerates to ordinary Karlsson scanning.) For traffic measurements one would probably prefer to keep a count for each level so that after the measurement we are able to draw a histogram showing the bandwidth distribution. For charging purposes we only need one counter, which we then increment with $c(k)$, where c is a charging function and k is the sample number $k \in \{1, 2, \dots, N\}$. A simple proposal for $c(k)$ would be $c(k) = k_1 * j + k_2$, where $j \in \{0, 1, \dots, L\}$ ($j = 0$ corresponds to bandwidth = 0) is a level-number and k_1 and k_2 are integer constants with $k_1 = 1$ and $k_2 = 0$ as the obvious choices. Some administrations might want to have a higher price penalty for high bandwidth utilisation and it might be desirable to have $k_2 > 0$ in order to give the customer bigger incentive to close down the connection, if it is not actually used over a longer period of time.

Let us now turn to the sampling method. First of all we need a measurement device that is synchronised to the access line in such a way that it is able to detect the individual cells and it has to know the protocol on the line in order to be able to detect dummy cells that do not carry any information. The measurement itself has to be a mean value taken over some time = T_s. If we have a non-linear charging function c, we need to keep track of each individual source because if high bandwidth utilisation is penalised via c it has to be on a per source basis. But even if $c(k) = j$ is used as a charging function as proposed above, we still need a counter per terminal, because the individual calls may have different destinations and will therefore have different scanning times. Anyway we choose the charging function $c(k) = j$, where $j \in \{0, 1, \dots, L\}$ for each sample, k, in the measurement, because it is the simplest one.

The minimum time we can have for T_s is the time it takes for a cell to pass the measurement mechanism = T_c. If the subscriber loop has the access rate C and the cell length is b_c we get

$$T_c = b_c / C.$$

Taking $T_s = T_c$ we would only be able to measure the rates 0 or C, but such a choice would make it impossible to detect the individual traffic bursts. If we make T_s too long, we will measure the mean value all the time, and in this case we could just have applied the ordinary Karlsson scanning as described in Section 3. We want T_s to be so small that it can detect a little part of a burst (whatever "little" then means in this connection). Burst detection (and definition) is a problem in itself and is treated in [8]. We also want to have a value for T_s that is independent of the policing that the sources might be subjected to because it is still under discussion if source policing should be performed or not with e.g. Roberts [4] representing the opposition and Rasmussen & Sørensen [6] the defense. A fixed time corresponding to a fixed number of cells for the averaging is preferable for implementation reasons. This number has to be an integer (to make it simple) constant times the number of levels, L, we want to distinguish. This integer represents (in T_c units) the time over which the sample should be taken. Somewhat arbitrarily we choose

8213 Computer N orks and ISDN Systems
20(1990)Decembe o.1/5, Amsterdam, NL

the constant to be 20 (what the real value needed should be will have to depend on experiments with real sources when they will be known)

$$T_s = b_c \cdot 20L/C.$$

We have now determined the duration of the "sample" we want to take. We now need to determine the time between two neighbouring samples (the h in Section 3). In [3] it is shown that the distance between the samples does not affect the mean value of the measurement but only the accuracy. From the sampling theory we know that all information outside the frequency band $[-\Omega, \Omega]$ (where the sampling period is $h = \pi/\Omega$) is lost, so there are limits to what we can do. The scanning interval (or sampling frequency) then will depend on the bandwidth function and on the application, because if we have a measurement we might want more detailed information than we need for charging purposes. Finally, since the charge is proportional to the sum of the c's, administrations will find it appropriate to let the call destination be a parameter that influences the sampling period. The smallest possible value for h is $h = T_s$, because having a smaller h would mean that we take a new sample point before we have finished measuring the previous one. The final determination of the sampling frequencies needed for various destinations would probably be done via a field trial. But in order to maintain the idea of scanning rather than sampling we shall take $h \gg T_s$.

5.2. Mathematical Model

For the assessment of the total charge c we sample the bandwidth function $B(t)$ in Fig. 1, where the kth sample has value j, with $j \in \{0, 1, \ldots, L\}$, where $j = 0$ corresponds to $w < \frac{1}{2}C/L$, $j = 1$ to $\frac{1}{2}C/L \leqslant w < \frac{3}{2}C/L$, $j = 2$ to $\frac{3}{2}C/L \leqslant w < \frac{5}{2}C/L$, ... and $j = L$ to $w \geqslant \frac{1}{2}(2L-1)C/L$. The following model is proposed: We regard j as a random variable whose state space is $\{0, 1, \ldots, L\}$. If the values of j occur with the probabilities

$$p_0, p_1, \ldots, p_L \quad \text{where} \sum_{j=0}^{L} p_j = 1,$$

we can assume, that when we have N samples the distribution will be $(L+1)$-nomial, i.e. if $x_0 + x_1 + \cdots + x_L = N$, then

$$f(x_0, x_1, \ldots, x_L) = \frac{N!}{x_0! x_1! \ldots x_L!} p_0^{x_0} p_1^{x_1} \cdots p_L^{x_L}$$

where as usual 0! is to be interpreted as 1. This function seems to be the least restrictive we can use to describe the sample values of the bandwidth function.

For this $(L+1)$-nomial distribution we have the expectation vector

$$E(x_0, x_1, \ldots, x_L) = \begin{bmatrix} E(x_0) \\ \vdots \\ E(x_L) \end{bmatrix} = \begin{bmatrix} Np_0 \\ \vdots \\ Np_L \end{bmatrix}$$

From this we can calculate the expectation of the charge first in terms of charging units assuming that the call received N scans

$$\mu = E[0 \cdot x_0 + 1 \cdot x_1 + \cdots + L \cdot x_L] = N \sum_{j=0}^{L} jp_j$$

In terms of "time-bandwidth-units" (as we have in Fig. 1) we have according to the translation rule given above

$$\mu = N \frac{hC}{L} \sum_{j=1}^{L} jp_j.$$

The variance can also be calculated, but it does not seem so interesting in this context.

8213 Comput etworks and ISDN Systems
20(1990)December, No.1/5, Amsterdam, NL

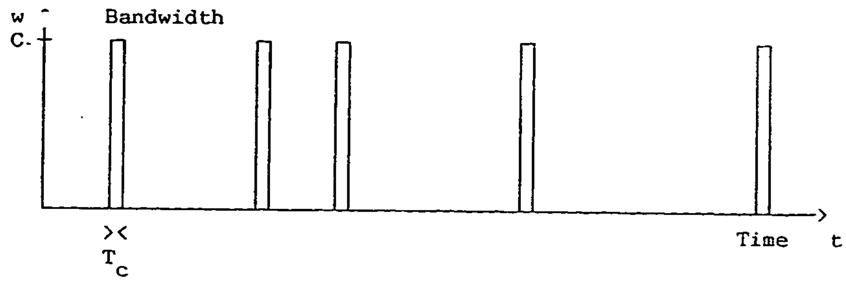

Fig. 2. Micro-view of bandwidth function.

5.3. Analysis of the Model

For the length of the call (T_2) we can still apply the analysis in Section 3 and hence we can conclude. that the number of scans N during a call, when considered as a random variable, has the correct mean value, viz. the same as the mean value of the distribution function describing T_2.

Let us now look at Fig. 1 again. If we call the bandwidth function $B(t)$, we want to prove, that $\int_0^{T_2} B(t)\,dt$ "on the average" is the same as the value μ calculated above (in the time bandwidth unit version), where $B(t)$ is to be interpreted as the outcome of an experiment so that we actually consider a whole family of B-functions where our B just is a sample (in the statistical sense) from this family.

Before we begin to assess the value of this integral, some comments on the possible members in the B-function family are in place. If we take a microscopic look at the B-functions, the appearance is very different from the kind of picture drawn in Fig. 1, because either the source emits a cell at the moment in which case it is the sole user of the whole bandwidth C available, or it does not emit a cell in which case the bandwidth is 0 (see Fig. 2).

This picture is disrupted by the measurement method applied so the way we regard Fig. 1 is as a picture taken from a long distance of a series of measurements where $h = T_s$ has been chosen. Such a picture might look as in Fig. 3, and this is actually the limit to which we can go (for physical reasons) evaluating the integral. It is clear that the integral of such a function will exist as long as it is taken over a finite interval. In the expectations calculated in Section 5.2 the x_i's represent the number of occurrences of the event $w = iC/L$. The j-factors in the μ calculation represent the number of area units for each count in the jth column and then the μ calculated corresponds to the area under the step-function shown in Fig. 3. But in Fig. 3 we have the limit situation where we have done the measurement as fine as possible and with the

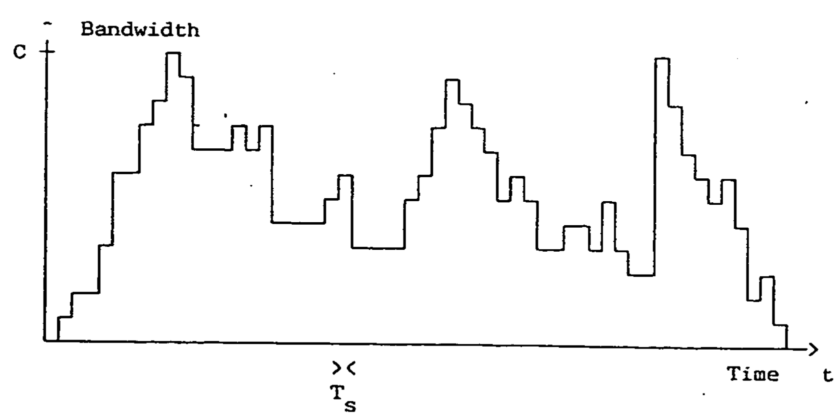

Fig. 3. Micro measurement.

8213 Computer Networks and ISDN Systems
20(1990)December No.1/5, Amsterdam, NL

highest possible sampling frequency and therefore for this case the μ we got is actually the same as the integral (by definition) and the proof is finished.

5.4. Example

In the CCITT SG XVIII [9] they have recently agreed that the cell length in ATM shall be 53 bytes with 48 bytes information and a 5-bytes header. If we take a subscriber access line with the rate $C = 34$ Mb/s, we get for the time it takes for a cell to pass the measurement mechanism

$$T_c = 53 \cdot 8/34 \cdot 10^6 \text{ s} = 12.5 \ \mu\text{s}$$

If we take the desired number of levels $L = 8$, we get for the duration of a sample

$$T_s = T_c \cdot 20L = 12.5 \cdot 20 \cdot 8 \ \mu\text{s} = 2.00 \text{ ms}.$$

Normally examples are used to verify, that the theory provided actually works when it comes to practical applications. As the only VBR-sources we know, with a reasonable amount of confidence, are voice sources, these will be used. Sriram and Whitt [10] report the following values for packetized voice traffic:

mean talkspurt duration:	352 ms,
mean silence duration:	650 ms (in active mode),
ADPCM-coding at bit rate	32 kb/s.

The probability to hit a 32 kb/s signal on a 34 Mb/s line is only $32 \cdot 10^3/34 \cdot 10^6 = 0.94 \cdot 10^{-3}$, but then the source is only active a $352/(352 + 650)$th of the time, i.e. 0.35, and the total probability to hit the signal active is only $0.33 \cdot 10^{-3}$, i.e. the probability not to hit is $1 - 0.33 \cdot 10^{-3}$, and since Sriram and Whitt assume geometric distribution in the active mode and exponential distribution for the pauses, we can assume the memoryless property, i.e. the probability not to hit the signal 160 consecutive times is $(1 - 0.33 \cdot 10^{-3})^{160} = 0.95$. I.e. the probability to hit at least one cell is 0.05, which does not seem satisfactory. Of course there is no reason to provide ATM just for voice sources, because the current circuit switched network does that in a cost effective manner already, so to provide a good example better data is needed.

6. Conclusion

This paper proposed a measurement method for VBR sources as they appear in ATM networks. The method is a generalisation of the well-known Karlsson scanning principle which is used for charging and measurement purposes in the Nordic countries and Germany. A proof is given which shows that the method on the average gives the correct value. In the end an example shows that the method is not very well suited to voice sources, but this is to be expected as ATM networks will not be provided just to satisfy the need for voice communication. Future studies of more realistic VBR sources are needed for a better assessment of the procedure proposed.

References

[1] S.A. Karlsson, Tekniska anordningar för samtaldebitering enligt tid (Technical devices for metering calls according to time, in Swedish and Finnish) *Helsingfors Telefonforening tekniska meddelande 1937* (2) (1937) 32–48.

[2] D. Henriet, Externalité d'encombrement et tarification au trafic: approche théorique des justifications et de la vulnérabilité par la concurrence (Congestion externality and traffic tarification: a theoretical approach of justifications and sustainability in a competitive environment, in French), *Ann. de Télécomm.* 42 (1–2) (1987) 13–19.

8213 Comput[illegible]etworks and ISDN Systems
20(1990)December, No.1/5, Amsterdam, NL

[3] V.B. Iversen, On general point processes in teletraffic theory with application to measurements and simulation, in: *Proc. ITC-8*, Melbourne (1976).
[4] J.W. Roberts, On flow control and admission control in an ATM-based multiservice network, Paper internal to the European project COST 224, Doc. no. 224 (89)-5.
[5] L. Mejlbro, Fouriertransformationen i L^2 og samplingssætningen, (The Fourier transform in L^2 and the sampling theorem, in Danish), Mathematical Institute, Technical University of Denmark, 1981.
[6] C. Rasmussen and J. Sørensen, Simple call acceptance procedures in an ATM network, ITC Specialist Seminar, Adelaide (1989).
[7] J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering* (Wiley, New York, 1965).
[8] L. Mejlbro and A.B. Nielsen, Burst detection, in: *Proc. 8th Nordic Teletraffic Seminar*, Helsinki (1989).
[9] CCITT SG XVIII, Draft recommendation I.361, Geneva, Feb. 1990.
[10] K. Sriram and W. Whitt, Characterizing superposition arrival processes in packet multiplexers for voice and data, *IEEE J. Select. Areas Comm.* 4 (6) (1986) 833–846.

Appendix 1. Translation into English of the chapter on the sampling theorem in [5]

It follows from the definition of an entire function, the definition of the exponential type and from Paley–Wieners' theorem that $L^2([-\pi, \pi])$ is a bijection onto the set of entire functions $f(z)$ of exponential type π for which

$$\int_{-\infty}^{+\infty} |f(x)|^2 \, dx < +\infty.$$

(for a more detailed explanation of these concepts the reader is referred to any standard work in the theory of complex functions (like [A1]). The set of all these entire analytic functions is called the *Paley–Wiener space* and is denoted $\boldsymbol{P}$, i.e.

$$\boldsymbol{P} = \left\{ f(z) \text{ entire} \,\middle|\, \int_{-\infty}^{+\infty} |f(x)|^2 \, dx < +\infty, \text{ and there is a } C > 0 \text{ such that for all } z \in \boldsymbol{C}: \; |f(z)| \leqslant C\, e^{\pi |z|} \right\}.$$

It is obvious, that $\boldsymbol{P}$ becomes a vector space over $\boldsymbol{C}$ using pointwise addition and scalar multiplication.

Since the restriction of $f \in \boldsymbol{P}$ to $\boldsymbol{R}$ belongs to $L^2(\boldsymbol{R})$, we have a natural inner product in $\boldsymbol{P}$ given by

$$\langle f, g \rangle_P = \langle f, g \rangle = \int_{-\infty}^{+\infty} f(x)\overline{g(x)} \, dx, \quad f(z), g(z) \in \boldsymbol{P}. \tag{A.1}$$

As the restriction of the functions $\boldsymbol{P}$ to $\boldsymbol{R}$ through the Fourier transform is a bijection onto the functions in $L^2([-\pi, \pi])$ and as the Fourier transform is an isomorphism, *$\boldsymbol{P}$ is a separable Hilbert space which is isometricly isomorphic with $L^2([-\pi, \pi])$.*

If $f \in \boldsymbol{P}$ has the representation

$$f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} \phi(t)\, e^{izt} \, dt,$$

where $\phi \in L^2([-\pi, \pi])$, it follows from Plancherel's theorem (see [A1] again) that

$$\| f \|_P^2 = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx = \int_{-\pi}^{+\pi} |\phi(t)|^2 dt = \| \phi \|_2^2. \tag{A.2}$$

As the Fourier coefficients for $\phi \in L^2([-\pi, \pi])$ with respect to the orthonormal base

$$\left\{ \frac{1}{\sqrt{2\pi}} e^{-int} \,\middle|\, n \in \boldsymbol{Z} \right\} \quad \text{are} \quad f(n) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} \phi(t)\, e^{int} \, dt,$$

8213 Computer Networks and ISDN Systems
20(1990)December No.1/5, Amsterdam, NL

we have (where the series is convergent in the L^2 sense)

$$\phi(t)=\sum_{n=-\infty}^{+\infty} f(n)\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-int},$$

and it follows from Parseval's theorem, that

$$\|\phi\|_2^2=\sum_{n=-\infty}^{+\infty} |f(n)|^2.$$

Hence, using (A.2) we get

$$\|f\|_P^2=\sum_{n=-\infty}^{+\infty} |f(n)|^2. \tag{A.3}$$

In particular we have $(f(n))_{n\in Z}\in l^2(Z)$.

As the inverse Fourier transform of $\mathrm{e}^{-int}\chi_{[-\pi,\pi]}/\sqrt{2\pi}$ is given by

$$\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{+\pi}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-int}\,\mathrm{e}^{ixt}\,\mathrm{d}t=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\mathrm{e}^{i(x-n)t}\,\mathrm{d}t=\frac{1}{2\pi}\left[\frac{\mathrm{e}^{i(x-n)t}}{i(x-n)}\right]_{t=-\pi}^{\pi}$$
$$=\frac{1}{\pi(x-n)}\frac{1}{2i}\{\mathrm{e}^{i(x-n)\pi}-\mathrm{e}^{-(x-n)\pi}\}=\frac{\sin[(x-n)\pi]}{(x-n)\pi}=(-1)^n\frac{\sin x\pi}{\pi(x-n)},$$

we conclude that the orthonormal base $\{(1/\sqrt{2\pi})\,\mathrm{e}^{-int}\mid n\in Z\}$ by the above mentioned isomorphism is mapped onto the orthonormal base $\{\sin[(z-n)\pi]/(z-n)\pi\mid n\in Z\}$ in P. Since $\phi(t)=\sum_{n=-\infty}^{+\infty} f(n)\,\mathrm{e}^{-int}/\sqrt{2\pi}$ by this isomorphism is mapped onto $f(z)$, we have

$$f(z)=\sum_{n=-\infty}^{+\infty} f(n)\frac{\sin[(z-n)\pi]}{\pi(z-n)}=\sum_{n=-\infty}^{+\infty}(-1)^n f(n)\frac{\sin\pi z}{\pi(z-n)}, \tag{A.4}$$

where the convergence of the series is defined by the metric in P. We shall make a closer examination of the convergence in P.

Let $g(z)=\sin\pi z/\pi z$, $z\in C\setminus\{0\}$, $g(0)=1$. Then $g(z)$ is an entire function. Due to the continuity there exists a $C>0$ such that

$$|g(z)|\leqslant C\quad\text{for } |x|\leqslant 1 \text{ and } |y|\leqslant 1,$$

where $z=x+iy$. Furthermore we have

$$|g(z)|\leqslant\frac{\mathrm{e}^{\pi|y|}}{\pi}\quad\text{for } |x|\leqslant 1 \text{ and } |y|\leqslant 1$$

and

$$|g(z)|\leqslant\frac{\mathrm{e}^{\pi|y|}}{\pi|y|}\quad\text{for } |x|\geqslant 1.$$

Now let $n_0\in Z$ be chosen so that $|x-n_0|\leqslant\frac{1}{2}$, and let $x_0=|x-n_0|$. If $|y|\leqslant 1$ we have the estimates

$$\sum_{n=-\infty}^{+\infty}\left|\frac{\sin[\pi(z-n)]}{\pi(z-n)}\right|^2=\sum_{n=-\infty}^{+\infty}|g(z-n)|^2=\sum_{n=-\infty}^{+\infty}|g(z-n_0-n)|^2$$
$$\leqslant 2C^2+2\sum_{n=-\infty}^{+\infty}\frac{\mathrm{e}^{2\pi|y|}}{\pi^2(n-x_0)^2}\leqslant 2C^2+\frac{2}{\pi^2}\mathrm{e}^{2\pi}\sum_{n=1}^{+\infty}\frac{1}{(n-\frac{1}{2})^2}<+\infty,$$

and if $|y|\geqslant 1$, we have the estimates

$$\sum_{n=-\infty}^{+\infty}\left|\frac{\sin[\pi(z-n)]}{\pi(z-n)}\right|^2=\sum_{n=-\infty}^{+\infty}|g(z-n_0-n)|^2\leqslant\mathrm{e}^{2\pi|y|}\left\{\frac{2}{\pi^2}+\frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{(n-\frac{1}{2})^2}\right\}<\infty.$$

8213 Comput[illegible] Networks and ISDN Systems
20(1990)December, No.1/5, Amsterdam, NL

These assessments show, that if $|y| \leqslant k$, then the sum $\sum_{n=-\infty}^{+\infty} |\sin[\pi(z-n)]/\pi(z-n)|^2$ is uniformly convergent.

According to Cauchy–Schwarz' inequality we have

$$\left| \sum_{n=-\infty}^{+\infty} f(n) \frac{\sin[\pi(z-n)]}{\pi(z-n)} \right|^2 \leqslant \sum_{n=-\infty}^{+\infty} |f(n)|^2 \sum_{n=-\infty}^{+\infty} \left| \frac{\sin[\pi(z-n)]}{\pi(z-n)} \right|^2 < +\infty.$$

hence we conclude that if $(f(n))_{n \in Z} \in l^2(Z)$, is given, the series

$$\sum_{n=-\infty}^{+\infty} f(n) \frac{\sin[\pi(z-n)]}{\pi(z-n)}$$

is absolutely and uniformly convergent in z, when $|\operatorname{Im} z| \leqslant k$. From this we conclude, that the convergence in $\boldsymbol{P}$ means that

$$f(z) = \sum_{n=-\infty}^{+\infty} f(n) \frac{\sin[\pi(z-n)]}{\pi(z-n)}$$

is uniformly convergent in every strip of the form $\{z \in C \mid |\operatorname{Im} z| < k\}$. Hence

$$f(z) = \sum_{n=-\infty}^{+\infty} (-1)^n f(n) \frac{\sin \pi z}{\pi(z-n)} \quad \text{for } z \in C \tag{A.5}$$

This result shows that every function $f \in \boldsymbol{P}$ can be reconstructed from the knowledge of the values $\{f(n) \mid n \in Z\}$. The series representation (A.5) is called the *cardinal series for f*. It was introduced for the first time in 1915 by Whittaker [A2]. Today this is known as the *sampling theorem*, due to the role it plays in communication theory (Shannon [A3] and Shannon & Weaver [A4]).

As a consequence of the above we have the following:

Theorem A.1. *To every complex series* $(a_n)_{n \in Z} \in l^2(Z)$ *there exists a uniquely determined element* $f \in \boldsymbol{P}$, *such that* $f(n) = a_n$. *This element is given by*

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n \frac{\sin[\pi(z-n)]}{\pi(z-n)} = \sum_{n=-\infty}^{+\infty} (-1)^n a_n \frac{\sin \pi z}{\pi(z-n)}, \quad z \in C$$

and $f(x) \in L^2(\boldsymbol{R}) \cap C^\infty(\boldsymbol{R})$ *for* $z = x$ *real.*

When this is translated into the argot used in the field of information theory the theorem is formulated:

Theorem A.2. *If the spectrum* $\phi(\omega)$ *of a function of the time* $f(t)$ *only contains frequencies from a finite band* $\omega \in [-\Omega, \Omega]$, *i.e.* $\phi(w) = 0$ *for* $|\omega| > \Omega$, *then* $f(t)$ *is completely determined by the values at the sampling points* $f(nT)$, $n \in Z$, *where the sampling period is* $T = \pi/\Omega$ *and*

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT) \frac{\sin[\Omega(t-nT)]}{\Omega(t-nT)}, \quad t \in \boldsymbol{R}.$$

Apart from the terminology, the only change here is that we consider the interval $[-\Omega, \Omega]$ instead of $[-\pi, \pi]$. It is assumed, of course, that $\phi \in L^2([-\Omega, \Omega])$.

In spite of the great usability of this result a few warnings should be given here. From a mathematical point of view it is absurd that a time dependent function is uniquely determined by the values at the sampling points. The theory only works because $f(t)$ implicitly is assumed to be analytic, i.e. $f(t)$ can be extended to an entire function. In particular $f(t)$ has a convergent series representation

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n t^n, \quad t \in \boldsymbol{R}.$$

8213 Computer Networks and ISDN Systems
20(1990)Decembe[illegible]p.1/5, Amsterdam, NL

But then the technique becomes absurd, because there is no reason whatsoever to assume that a physical or technical process can be described by an analytical function. On the contrary there is every reason to believe that this cannot be done by a C^∞-function. As long as we consider the whole real axis $\boldsymbol{R}$, the analytical functions, whose domain contains $\boldsymbol{R}$, lie extremely sparse in $C^\infty(\boldsymbol{R})$. (Just imagine, that an analytical function of the kind considered is *uniquely* determined all over its domain, when one only knows the values in an arbitrary small interval $]-\epsilon, \epsilon[$.) Hence there is every reason to believe, that the assumption $\phi(\omega)=0$ for almost all ω outside $[-\Omega, \Omega]$ is *never* fulfilled.

Well, one would say, why does it work anyway? The secret is that in practice one is only able to measure $f(nT)$ for a *finite* number of sampling points, and one puts $f(nT)=0$ for $|n|>N$, which corresponds to a consideration of the time function $f(t)$ only in the finite time interval $[-NT, NT]$, and in each such compact interval Weierstrass' approximation theorem already shows, that every continuous function can be approximated uniformly by using analytical functions (even by polynomials!), and the absurdity has disappeared.

As it was assumed above that $f(nT)=0$ for $|n|>N$ and implicitly that $f(nT)<+\infty$ for $|n|\leqslant N$, the numbers $a_n=f(nT)$ will define a series from $L^2(Z)$. Since we in any case for $(c_n)_{n\in Z}\in L^2(Z)$ shall have, that $c_n\to 0$ for $n\to+\infty$, and as the convergence of the series

$$\sum_{n=-\infty}^{+\infty} c_n \frac{\sin[\Omega(t-nT)]}{\Omega(t-nT)}$$

is absolute and uniform, we only need to choose N so that $\Sigma_{|n|>N}|c_n|^2<\epsilon^2$, in order to ensure that the error we commit by putting $a_n=0$ for $|n|>N$ and $a_n=c_n$ otherwise, is very small. (How this is done in practice must depend on intuition, as the c_n for $|n|>N$ are precisely the values that have not been measured.) It would be obvious to put $f(t)=0$ for $|t|\geqslant N+1$, so that $|f(t)|<\epsilon\cdot \mathrm{cl}$, where cl is a constant that depends only on the series $(\sin[\Omega(t-nT)]/\Omega(t-nT))_{n\in Z}$. This is equivalent to saying that we only consider the values of $f(t)$ for $t\in[-NT, NT]$. For if $f(t)\equiv 0$ for $|t|\geqslant N+1$, then we would have $f(z)\equiv 0$ for all $z\in Z$, since $f(z)$ was an entire function.

In practice one should not be confused by this. The theory of Complex Functions (and by the way also Functional Analysis) is a powerful tool in the solution of these problems even though nature does not always behave like an analytic function.

References

[A1] J.L. Marsden, *Basic Complex Analysis* (Freeman, San Francisco, 2nd ed., 1973).

[A2] E.T. Whittaker, On the functions which are represented by the expansions of the interpolation theory, *Proc. Roy. Soc. Edinburgh* 35 (1915) 181–194.

[A3] C.E. Shannon, Communication in the presence of noise, *Proc. IRE* 37 (1) (1949) 10–21.

[A4] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, 1949).